3D Motion Estimation of Human Head by Using Optical Flow

Ján MIHALÍK, Viktor MICHALČIN

Laboratory of Digital Image Processing and Videocommunications, Dep. of Electronics and Multimedia Telecommunications, FEI TU Košice, Park Komenského 13, 041 20 Košice, Slovak republic

Jan.Mihalik@tuke.sk, michalcin@vk.upjs.sk

Abstract. *The paper deals with the new algorithm of estimation of large 3D motion of the human head by using the optical flow and the model Candide. In the algorithm prediction of 3D motion parameters in a feedback loop and with multiple iterations was applied. The prediction of 3D motion parameters does not require creating of the synthesized frames but directly uses the frames of input videosequence. Next the algorithm does not need extracting of feature points inside the frames because they are given by the vertices of the used calibrated model Candide. As achieved experimental results show, the iteration process in prediction of 3D motion parameters increased the accuracy of estimation above all the large 3D motion. Such a way the estimation error is decreased without its accumulation in long videosequence. Finally the experimental results show that for 3 iterations a state of saturation was achieved what means that by next increasing of the number of iterations practically no significant increasing of the accuracy of estimation of 3D motion parameters is occurred.*

Keywords

Optical flow, 3D motion, estimation, prediction, modeling, human head, algorithm.

1. Introduction

Compression of classical standard videocodecs H.261, H.263, H.264, MPEG-1, MPEG-2, MPEG-4 [1] is based on reduction of the intra-frame and inter-frame redundancy of videosignals. Their core consists of the inter-frame hybrid coding system [2] with motion estimation and compensation. Disadvantage of the standard videocodecs is considerable loss of the visual quality of the output videosignal in case of very low bit rate (less than 64kb/s) coding.

If semantic information about the content of frames is known, very effective coding of the videosignal by model based video coding [3], [4] is possible. The coding is based on modeling of videoobjects inside of a visual scene by using three dimensional (3D) models. Each frame is analyzed in the coder to obtain parameters for 3D models. Obtained parameters express for example deformation or 3D motion of the object in the scene. Compared to the classical videocodecs in this case only the parameters are coded and transmitted instead all picture elements of the frames for the classical videocodecs. The result is very low bit rate in output of the coder.

Location of the human head object in the videosequence is given by its 3D global motion in each frame. 3D global motion is defined by six parameters, three rotation angles around all axes in 3D coordinate system and three translation components along all axes in this coordinate system. In general, the algorithms of estimation of 3D motion published in scientific papers are divided into two groups. The first group is created from algorithms based on tracking of extracted feature points in frames [5], [6]. The second group is composed from the algorithms based on estimation by the optical flow [7], [8].

The algorithms based on tracking of extracted feature points [9] assume extracted feature points from the human head in each frame. The accuracy of estimated 3D motion parameters depends mainly on the accuracy of extracted feature points in the frames. Further the accuracy depends on the number of extracted feature points of the human head used by estimation. Because almost all feature points are lying on edges in cases when only the profile of human head is shown in a frame there is a limited number of unambiguous extracted feature points. The small number of extracted feature points used by estimation can result in inaccuracy of the estimated 3D motion parameters.

Estimation based on the optical flow uses the optical flow equation, which describes the relationship between 2D motion parameters of the feature point and time-spatial derivatives of the image luminance in the point [10]. Then 3D motion parameters of the corresponding vertex of 3D wire frame model of the human head are estimated by using the optical flow equation. It is possible to use corresponding feature points from the whole human head region in the videosequence [11]. There is no need to extract all feature points in each frame. The feature points of the human head are determined on beginning simply by

using projection of the corresponding vertices of 3D wire frame model on the first frame of the input videosequence. Compared to the algorithm based on tracking of extracted feature points in each frame it is possible to use more feature points. At the maximum case they are given by all corresponding vertices of 3D wire frame model of the human head. From this it follows out that estimation by using the optical flow method is not so sensitive on inaccuracy of the extracted feature points. Our new algorithm of estimation of 3D motion parameters introduced in this paper is based on the optical flow equation.

2. Optical Flow

The optical flow can be defined as a field of 2D vectors (u_i,u_j). It describes the direction and magnitude of motion of all points in a frame of an input videosequence during the time interval Δt. Assume the point at the position (i,j) in the frame N and its moving to the new position (i',j') in the frame N+1 as it is seen in Fig.1. Then the motion of this point between two consecutive frames is defined by the two components u_i and u_j of the motion vector of optical flow.

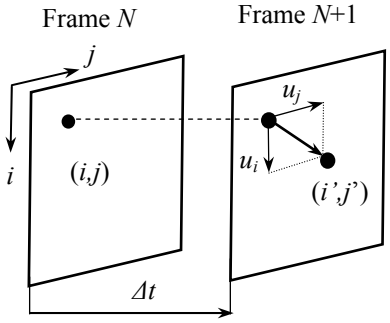


Fig. 1. Vector of the optical flow.

The optical flow is calculated from 3D luminance function $I(i,j,t)$. If a point is moved in the videosequence during the time interval Δt then its luminance value $I(i+\Delta i,j+\Delta j,t+\Delta t)$ will not change. Assuming constant light conditions in the visual scene, it means that the luminance of the moving point in the videosequence remains constant, i.e.

$$I(i,j,t) = I(i+\Delta i, j+\Delta j, t+\Delta t)\,. \tag{1}$$

The Taylor expansion of the right side of eq. (1) is

$$I(i+\Delta i, j+\Delta j, t+\Delta t) = I(i,j,t) + \frac{\partial I(i,j,t)}{\partial i}\Delta i + \frac{\partial I(i,j,t)}{\partial j}\Delta j + \frac{\partial I(i,j,t)}{\partial t}\Delta t + \varepsilon \tag{2}$$

where ε is the part with higher derivatives. Inserting (2) into (1) and after dividing by Δt we get

$$\frac{\partial I(i,j,t)}{\partial i}\frac{\Delta i}{\Delta t} + \frac{\partial I(i,j,t)}{\partial j}\frac{\Delta j}{\Delta t} + \frac{\partial I(i,j,t)}{\partial t} + \varepsilon(\Delta t) = 0 \tag{3}$$

where we assume that Δi and Δj are varying with Δt. If $\Delta t \to 0$ and the motion of the tracking point is smooth then the term $\varepsilon(\Delta t)$ can be disregarded and eq. (3) will be

$$\frac{\partial I(i,j,t)}{\partial i}\frac{di}{dt} + \frac{\partial I(i,j,t)}{\partial j}\frac{dj}{dt} + \frac{\partial I(i,j,t)}{\partial t} = 0\,. \tag{4}$$

Let in eq. (4)

$$\frac{di}{dt} = u_i'\,, \tag{5}$$

$$\frac{dj}{dt} = u_j'\,, \tag{6}$$

then we get the equation of optical flow

$$I_i u_i' + I_j u_j' + I_t = 0 \tag{7}$$

where I_i, I_j are the partial spatial derivatives of the luminance function $I(i,j,t)$ with respect i, j and I_t is the partial derivative with respect t. The derivatives I_i, I_j, I_t can be determined from the input videosequence. For practical purposes it is better to multiply (7) by dt to get it in the form

$$I_i u_i + I_j u_j + \Delta I_t = 0 \tag{8}$$

where $u_i = i'-i$ and $u_j = j'-j$ are the components of the vector of optical flow and $\Delta I_t = I(i,j,t+1) - I(i,j,t)$ is the difference of luminance function $I(i,j,t)$ in the time direction. Assuming that near points move together, a system of linear equations with two unknown components (u_i,u_j) of the vector of optical flow can be obtained from (8).

The partial derivatives I_i,I_j and the time luminance difference ΔI_t were approximated by the numerical methods using the frames of the input videosequence in the window of the size 3x3 points. Approximation of I_i,I_j was done by Sobel operator [12] in the point (i,j,t) of the frame N and ΔI_t was calculated as an average of the time luminance differences between the frames N and N+1 in the same window.

3. Small 3D Motion Estimation

In this section we derive equations for small motion estimation by using the optical flow between the two successive frames where 3D wire frame model of a human head and the camera were calibrated by the first (reference) frame [15]. Assume that the vertex $(h,v,r)^T$ of the calibrated 3D wire frame model in the model coordinate system (MCS) is in the initial position. For it we determine its corresponding point (i,j) by the perspective projection on the reference frame. In case of known 3D motion parameters we can calculate the new position of the vertex $(h',v',r')^T$ in MCS or converted to (x',y',z') in the camera coordinate system (CCS) by 3D motion equation [13]

$$x' = x\left[1 + \frac{1}{x}\left(-\Theta_r y + \Theta_v (d-z) + t_h\right)\right], \tag{9}$$

$$y' = y\left[1+\frac{1}{y}\left(\Theta_r x-\Theta_h(d-z)+t_v\right)\right], \tag{10}$$

$$z' = z\left[1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)\right]. \tag{11}$$

Dividing (9) and (10) by (11) we have

$$\frac{x'}{z'} = \frac{x}{z}\frac{\left[1+\frac{1}{x}\left(-\Theta_r y+\Theta_v(d-z)+t_h\right)\right]}{\left[1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)\right]}, \tag{12}$$

$$\frac{y'}{z'} = \frac{y}{z}\frac{\left[1+\frac{1}{y}\left(\Theta_r x-\Theta_h(d-z)+t_v\right)\right]}{\left[1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)\right]}. \tag{13}$$

For the corresponding point (*i'*,*j'*) in the successive frame we get from (12) and (13) by using the perspective projection equation [13]

$$(j'-j) = \frac{j-j_0}{x}\left(-\Theta_r y+\Theta_v(d-z)+t_h\right) - \frac{j'-j_0}{z}\left(\Theta_v x-\Theta_h y-t_r\right), \tag{14}$$

$$(i'-i) = \frac{i-i_0}{y}\left(\Theta_r x-\Theta_h(d-z)+t_v\right) - \frac{i'-i_0}{z}\left(\Theta_v x-\Theta_h y-t_r\right) \tag{15}$$

where the differences (*i'-i*) and (*j'-j*) are the components (u_i,u_j) of the vector of optical flow in Fig. 1. After substitution u_i and u_j in (14) and (15) will be

$$u_j = \frac{j-j_0}{x}\left(-\Theta_r y+\Theta_v(d-z)+t_h\right) - \frac{j'-j_0}{z}\left(\Theta_v x-\Theta_h y-t_r\right), \tag{16}$$

$$u_i = \frac{i-i_0}{y}\left(\Theta_r x-\Theta_h(d-z)+t_v\right) - \frac{i'-i_0}{z}\left(\Theta_v x-\Theta_h y-t_r\right). \tag{17}$$

Let in (16), (17)

$$j'-j_0 = j'-j_0+j-j = (j'-j)+(j-j_0) = u_j+(j-j_0), \tag{18}$$

$$i'-i_0 = i'-i_0+i-i = (i'-i)+(i-i_0) = u_i+(i-i_0), \tag{19}$$

then for the components (u_i,u_j) of the vector of optical flow we have

$$u_j = \frac{\frac{j-j_0}{x}\left(-\Theta_r y+\Theta_v(d-z)+t_h\right)}{1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)} - \frac{\frac{j-j_0}{z}\left(\Theta_v x-\Theta_h y-t_r\right)}{1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)}, \tag{20}$$

$$u_i = \frac{\frac{i-i_0}{y}\left(\Theta_r x-\Theta_h(d-z)+t_v\right)}{1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)} - \frac{\frac{i-i_0}{z}\left(\Theta_v x-\Theta_h y-t_r\right)}{1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right)}. \tag{21}$$

From previous equations the nonlinear dependence of the components (u_i,u_j) on 3D motion parameters Θ_h, Θ_v, Θ_r, t_h, t_v, t_r follows out. A solution of the nonlinear system is complex and needs a lot of operations. Assuming small rotation angles (Θ<<1) and the large distance *d* between the camera and the human head compared to the depth coordinate *r* (*z*=*d*-*r*) of the vertices of 3D model, for the denominators in (20) and (21) the following simplification is valid

$$1+\frac{1}{z}\left(\Theta_v x-\Theta_h y-t_r\right) \cong 1. \tag{22}$$

Referred to (22) and by using the perspective projection equation we get u_i and u_j both linearly depending on 3D motion parameters $\overline{\mathbf{P}} = (\Theta_h,\Theta_v,\Theta_r,t_h,t_v,t_r)^T$

$$u_j = -\frac{JI}{f_y}\Theta_h + \left(\frac{f_x r}{(d-r)}-\frac{J^2}{f_x}\right)\Theta_v + \frac{f_x I}{f_y}\Theta_r + \frac{f_x}{(d-r)}t_h + 0t_v + \frac{J}{(d-r)}t_r = \overline{\mathbf{V}}\ \overline{\mathbf{P}}, \tag{23}$$

$$u_i = \left(\frac{f_y r}{(d-r)}-\frac{I^2}{f_y}\right)\Theta_h - \frac{JI}{f_x}\Theta_v - \frac{f_y J}{f_x}\Theta_r + 0t_h - \frac{f_y}{(d-r)}t_v + \frac{I}{(d-r)}t_r = \overline{\mathbf{U}}\ \overline{\mathbf{P}} \tag{24}$$

where I=(i-i_0), J=(j-j_0) are the centered coordinates of the point in the initial frame and $\overline{\mathbf{U}},\overline{\mathbf{V}}$ are the line vectors for simplification of both equations. After substitution of (23) and (24) into (8) we have the linear equation

$$(I_i\overline{\mathbf{U}}+I_j\overline{\mathbf{V}})\overline{\mathbf{P}} = -\Delta I_t \tag{25}$$

for one vertex of 3D model in MCS and its perspective projected point in the frame. For exact computation of 3D

small motion parameters $\overline{\mathbf{P}}$ it is needed to use 6 eq. (25) for 6 vertices of 3D model. Next the vertices will be named the feature vertices and their perspective projected points in the frame the feature points. For more accurate estimation we have to use more then 6 feature vertices. Then we have the system of linear equations

$$\underline{\mathbf{Z}}\overline{\mathbf{P}} = -\Delta\overline{\mathbf{I}}_t \tag{26}$$

where the separate lines of the matrix $\underline{\mathbf{Z}}$ and the components of the vector $\Delta\overline{\mathbf{I}}_t$ on the right side are composed from (25). We used the least square method (LSM) for solution of the motion parameters from (26)

$$\overline{\mathbf{P}} = -(\underline{\mathbf{Z}}^T\underline{\mathbf{Z}})^{-1}\underline{\mathbf{Z}}^T\Delta\overline{\mathbf{I}}_t\,. \tag{27}$$

The most suitable feature vertices of 3D model of the human head for 3D motion estimation are vertices where only 3D global motion can be presented. Using of features vertices where 3D local motion is expected leads to inaccuracy of estimation.

4. Large 3D Motion Estimation

Often in the real videosequences 3D motion of the human head is large. Because of linearization of (4), (22) and also the equation of 3D motion [13], 3D motion parameters of the large 3D motion are estimated with a higher error.

In case of the large 3D motion estimation it is possible to use a prediction of 3D motion parameters from the previous frame as it is seen in Fig. 2. The parameters for the actual frame are predicted by the parameters $\hat{\overline{\mathbf{P}}} = (\hat{\Theta}_h, \hat{\Theta}_v, \hat{\Theta}_r, \hat{t}_h, \hat{t}_v, \hat{t}_r)$ from the previous frame.

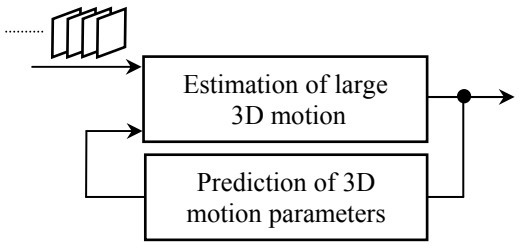


Fig. 2. Estimation of the large 3D motion by using of prediction of its parameters.

On the basis of (23) and (24) the prediction of the vector of optical flow is

$$\hat{u}_j = -\frac{JI}{f_y}\hat{\Theta}_h + \left(\frac{f_x r}{(d-r)} - \frac{J^2}{f_x}\right)\hat{\Theta}_v + \frac{f_x I}{f_y}\hat{\Theta}_r + \\ + \frac{f_x}{(d-r)}\hat{t}_h + 0\hat{t}_v + \frac{J}{(d-r)}\hat{t}_r, \tag{28}$$

$$\hat{u}_i = \left(\frac{f_y r}{(d-r)} - \frac{I^2}{f_y}\right)\hat{\Theta}_h - \frac{JI}{f_x}\hat{\Theta}_v - \frac{f_y J}{f_x}\hat{\Theta}_r + \\ + 0\hat{t}_h - \frac{f_y}{(d-r)}\hat{t}_v + \frac{I}{(d-r)}\hat{t}_r \tag{29}$$

where $\hat{u}_i, \hat{u}_j$ are the predicted components of the optical flow vector for the selected feature point in the actual frame and $\hat{\Theta}_h, \hat{\Theta}_v, \hat{\Theta}_r, \hat{t}_h, \hat{t}_v, \hat{t}_r$ are 3D motion parameters from the previous frame.

Assuming smooth motion of the human head in the videosequence the absolute prediction error $|(u_i - \hat{u}_i)|$ is smaller then the absolute value $|u_i|$. The same is valid for the component u_j. The knowledge is utilized to estimate the prediction error vector $\left((u_i - \hat{u}_i),(u_j - \hat{u}_j)\right)$ instead of the component vector (u_i, u_j) when the smaller linearization error is achieved. Then we can get more accurate estimation of 3D motion parameters.

On the basis of the above knowledge we derive the equations for estimation of the large 3D motion of the human head by using the optical flow between the two successive frames where the first frame is the reference one. Assume that the luminance of the moved point in the videosequence remains constant

$$I(i + u_i, j + u_j, t+1) = I(i, j, t)\,. \tag{30}$$

By inserting the predicted components $\hat{u}_i, \hat{u}_j$ to the left side of (30) we have

$$I\left(i + \hat{u}_i + (u_i - \hat{u}_i), j + \hat{u}_j + (u_j - \hat{u}_j), t+1\right) = I(i, j, t)\,. \tag{31}$$

After Taylor expansion of the left side of (31) and disregarding of the term with higher derivatives we get

$$I_i(i + \hat{u}_i, j + \hat{u}_j, t+1)(u_i - \hat{u}_i) + \\ + I_j(i + \hat{u}_i, j + \hat{u}_j, t+1)(u_j - \hat{u}_j) + \Delta\hat{I} = 0 \tag{32}$$

where $\Delta\hat{I} = I(i + \hat{u}_i, j + \hat{u}_j, t+1) - I(i, j, t)$.

Let $\hat{I}_i = I_i(i + \hat{u}_i, j + \hat{u}_j, t+1) = \partial I(i + \hat{u}_i, j + \hat{u}_j, t+1)/\partial i$ and $\hat{I}_j = I_j(i + \hat{u}_i, j + \hat{u}_j, t+1) = \partial I(i + \hat{u}_i, j + \hat{u}_j, t+1)/\partial j$, then by substitution and rearrangement in (32) we have

$$\hat{I}_i u_i + \hat{I}_j u_j = -\Delta\hat{I}_t \tag{33}$$

where $\Delta\hat{I}_t = \Delta\hat{I} - \hat{I}_i\hat{u}_i - \hat{I}_j\hat{u}_j$. In (33) there are the same two unknown components of the vector (u_i,u_j) like in (8). The difference between (8) and (33) is that the partial spatial derivatives $\hat{I}_i, \hat{I}_j$ in (33) are calculated in the actual

frame where the 3D motion is estimated while in (8) in the reference frame. After inserting (23), (24) into (33) we get the linear equation for one feature vertex of 3D model

$$(\hat{I}_i\overline{\mathbf{U}} + \hat{I}_j\overline{\mathbf{V}})\overline{\mathbf{P}} = -\Delta\hat{I}_t . \quad (34)$$

For accurate estimation of 3D motion parameters $\overline{\mathbf{P}}$ it is needed to select more than 6 feature vertices of 3D model. Then we have the system of linear equations which is similar to (26)

$$\hat{\underline{\mathbf{Z}}}\overline{\mathbf{P}} = -\Delta\hat{\underline{\mathbf{I}}}_t \quad (35)$$

and which we solve by using LSM as follows

$$\overline{\mathbf{P}} = -(\hat{\underline{\mathbf{Z}}}^T\hat{\underline{\mathbf{Z}}})^{-1}\hat{\underline{\mathbf{Z}}}^T\Delta\hat{\underline{\mathbf{I}}}_t . \quad (36)$$

An accuracy of the algorithm of 3D motion estimation based on (36) we can increase by using prediction of the motion parameters in the iterative feedback loop. The components of the vector of optical flow are then predicted from 3D motion parameters estimated in the previous iteration while estimation is running always between the reference and actual frame. By the beginning the prediction $\hat{u}_i^1, \hat{u}_j^1$ by using (28) and (29) is done by the components $u_i(t)$, $u_j(t)$ from the previous frame

$$\hat{u}_i^1(t+1) = u_i(t), \quad (37)$$

$$\hat{u}_j^1(t+1) = u_j(t). \quad (38)$$

In general for the next iterations when n=2,3,4,… we can write

$$\hat{u}_i^n(t+1) = u_i^{n-1}(t+1), \quad (39)$$

$$\hat{u}_j^n(t+1) = u_j^{n-1}(t+1). \quad (40)$$

3D motion estimation by using the optical flow is the fast method and we can consider it as a solution of the difference problem, because in each frame it is needed to determine the partial derivatives. Its complexity is given by the number of the selected feature points.

5. Experimental Results

Experimental results of 3D motion estimation of the human head by using the optical flow have been obtained for the testing videosequence "MissAmerica" with the frame rate 30Hz and the size 288x352 pels. As a specific 3D model of the human head we used 3D wire frame model Candide [14] which was calibrated by the first (reference) frame of the videosequence "MissAmerica" in Fig. 3a. For calibration of 3D model Candide we used the affine method [15] with the manual fitting correction. The result of this calibration is shown in Fig. 3b. Further for the purpose of obtaining the parameters of a camera *d, fx, fy* we calibrated the camera by using the reference frame and assuming zero 3D motion parameters [13]. For the distance d=400 pels we obtained the scaled focal lengths of the camera fx=354 pels and fy=333 pels.

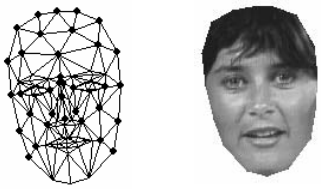
a) b)

Fig. 3. a) The first frame of videosequence „MissAmerica", b) the calibrated model Candide after projection on the first frame.

It is very difficult to use an objective criterion for direct measuring of the accuracy of estimated 3D motion parameters of the human head in the real videosequences, because their exact values are not known forehand. Therefore as the objective criterion for measuring of the accuracy of the estimated 3D motion parameters we use the peak signal/noise ratio (SNR) for the region of the human head in the frame

$$SNR = 10\log\frac{255^2}{\frac{1}{N}\sum_{(i,j)}\left[I_{orig}(i,j,t) - I_{synt}(i,j,t)\right]^2} \quad (41)$$

where $I_{orig}(i,j,t)$ is the luminance of the input frame, $I_{syn}(i,j,t)$ – synthesized frame and N is number of pels in the region of the human head in these frames. For the videosequence MissAmerica the number N was about 12000 pels. To compare the results of 3D motion estimation, we used the plane algorithm based on two dimensional affine transformation [13] in all experiments for texturing of the human head model. For illustration, in Fig. 4b an example of the textured model Candide in the synthesized frame is shown. Note that the choice of the algorithm of texturing has not any direct impact on the accuracy of 3D motion estimation. Then conclusions for 3D motion estimation in this paper are valid for using any algorithm of texturing [16]. The subjective evaluation of 3D model adaptation to the human head after its projection on the frames is very important. Therefore we evaluated the obtained results of our experiments by this criterion, too.



a) b)

Fig. 4. Model Candide a) with marked feature vertices for 3D motion estimation, b) textured by the plane algorithm.

First we estimated 3D motion parameters $\overline{\mathbf{P}}$ between two successive frames by using the algorithm for estimation the small 3D motion based on. (27) and then we used the algorithm for estimation large 3D motion based on (36). Exact calculation by using 6 feature vertices gives inaccurate results therefore we increased the number of the feature vertices on 35. All selected 35 vertices of the model Candide are shown in Fig. 4a. By the beginning of the estimation all 35 vertices are projected on the first (reference) frame for purpose to obtain the derivatives of the luminance function $I(i,j,t)$. The selection of the feature vertices was done with assumption that they or their corresponding points on the human head in the frame make only the global 3D motion. With this selection we eliminated a possible influence of the local 3D motion on the accuracy of estimation of 3D global motion parameters.

In Fig. 5 the graphs of SNR are shown for the first 35 frames for the algorithms of small 3D motion estimation based on (27), and large 3D motion estimation based on. (36). From these graphs it follows out that the algorithm of large 3D motion estimation gives better results and is more accurate what confirms our theoretical assumptions. SNR for the algorithm of large 3D motion estimation is higher in comparison to that one for the algorithm of small 3D motion estimation and in average it is about 2,54dB.

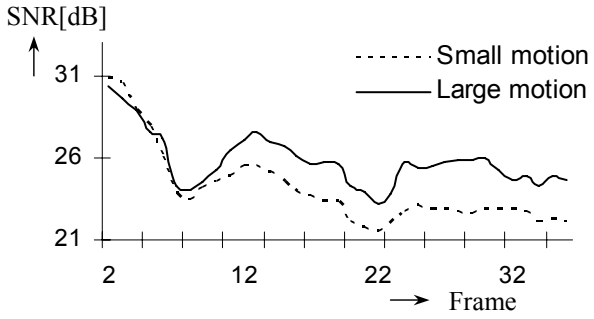


Fig. 5. The dependences of SNR of the first 35 frames for the algorithms of small and large 3D motion estimation.

By decimation of the videosequence MissAmerica in time we decreased its frame frequency on 15Hz assuming larger motion between two successive frames than in the videosequence with the frame frequency 30Hz. By using this decimated videosequence we discovered the effect of the iteration process described by (39) and (40) on minimizing of the estimation error for the algorithm of large 3D motion estimation. Fig. 6 shows the influence of the number of iterations on the accuracy of estimation of the parameters of large 3D motion based on (36) for the frame frequency 15 Hz. We used the iteration process described by (39) and (40) with 1, 2, 3 and 10 iterations. In case of the first iteration we do not reach the considerable improvement compared to the estimation of small motion in 30Hz videosequence and in average it is 1,08dB. In the videosequence with the frame frequency 15Hz the global 3D motion is too large and therefore it is not enough to use only the one iteration to achieve the same accuracy as for the videosequence with 30Hz. If we use two iterations in the algorithm of large 3D motion estimation SNR grows up in average about 3,27dB compared to that one of the algorithm of small 3D motion estimation and about 2,21dB compared to the result of the algorithm of large 3D motion estimation with one iteration. For three or more iterations there is not significant grooving of SNR what gives a saturation state as is shown in Fig. 7. It means that additional increasing of the number of iteration does not affect next increasing of the accuracy of the estimated 3D motion parameters.

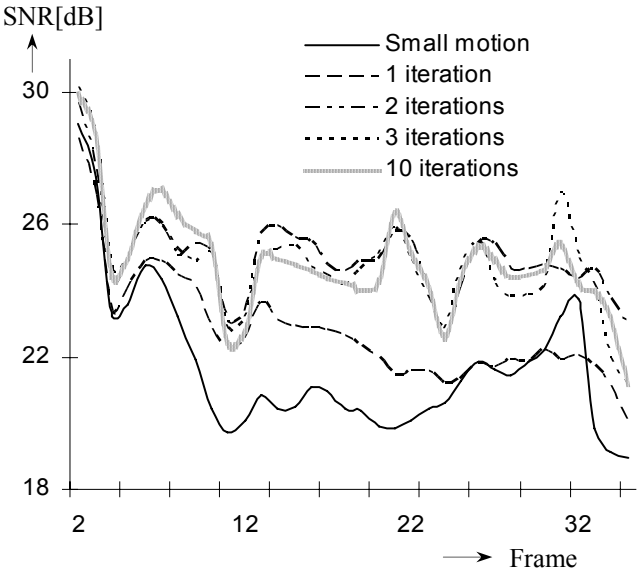


Fig. 6. Influence of the number of iterations on the accuracy of estimation of the large 3D motion parameters for the frame frequence 15 Hz.

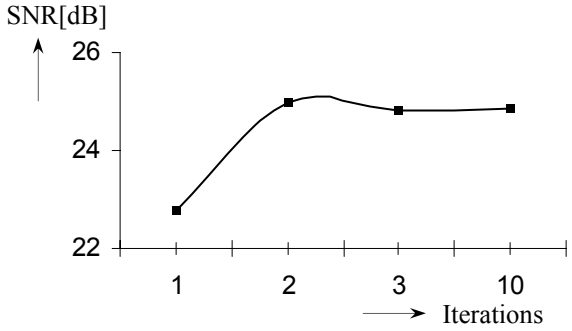


Fig. 7. Dependence of the average SNR on the number of iterations for the frame frequence 15 Hz.

For the subjective evaluation the frames of the number 3, 47 and 74 of the videosequence „MissAmerica“ with the frame frequency 15 Hz and with the projected model Candide after 3D motion estimation are shown in Fig. 8. For the algorithm of small 3D motion estimation (Fig. 8a) the influence of the estimation error and its accumulation is evident. The estimation error grows up in next frames and causes the distortion of the estimated 3D motion parameters $\Theta_h, \Theta_v, \Theta_r, t_h, t_v, t_r$ what results in the bad position of the moved 3D model Candide in MCS. The frames of the number 47 and 74 (Fig.8a) present mainly errors in the rotations angles. On the other side for the algorithm of large 3D motion estimation with one iteration (Fig. 8b)

the estimation error and its accumulation is not so much visible, but in the frames of the number 47 and 74 there are still small inaccuracies in fitting caused mainly by errors in the translation parameters. If we increase the number of iterations to 3 then we get very accurate estimation of 3D motion of the human head (Fig. 8c).

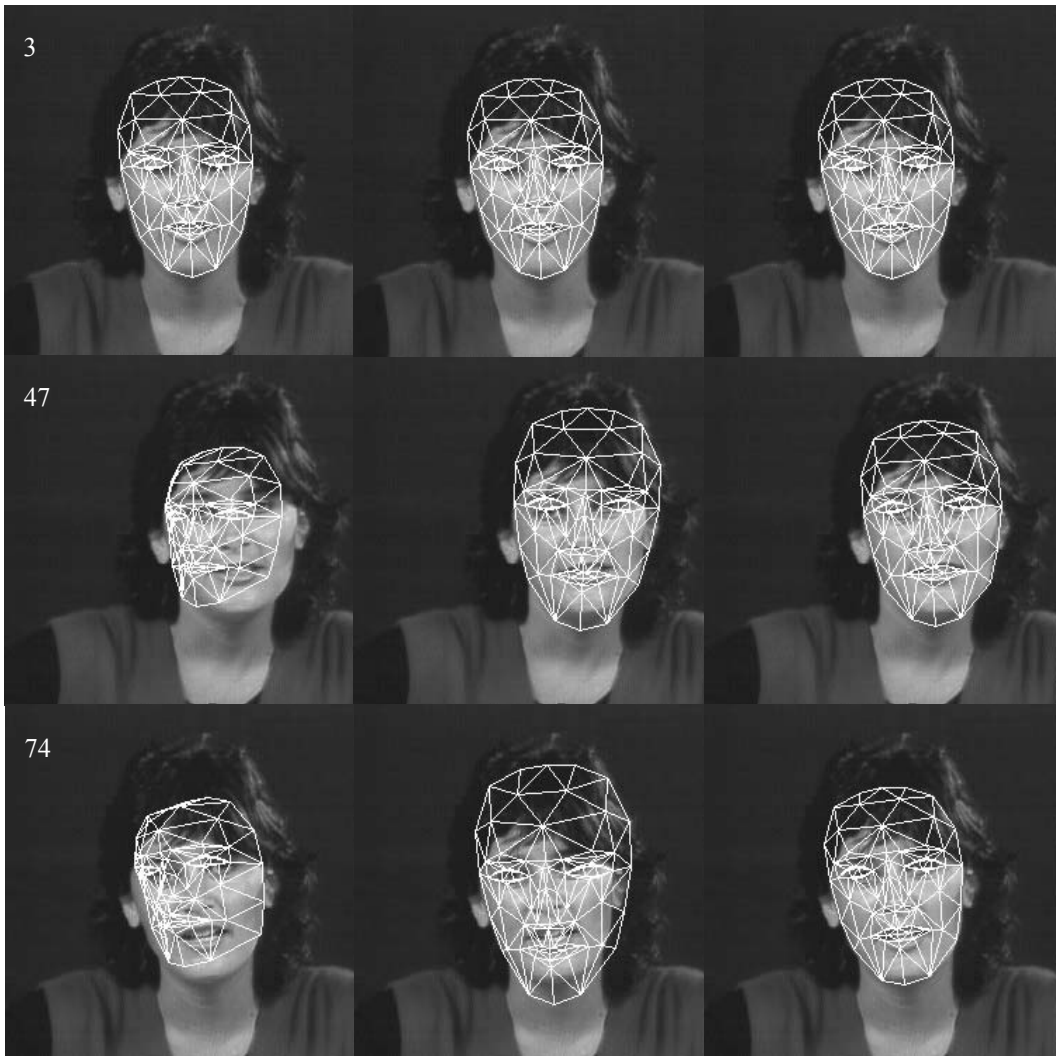


Fig. 8. Frames, the number 3, 47 and the last frame 74 of the videosequence „MissAmerica“ with the frame frequence 15 Hz and the projected model Candide after estimation a) small 3D motion, b) large 3D motion with 1 iteration, c) large 3D motion with 3 iterations.

6. Conclusion

The main subject of this paper was 3D motion estimation of the human head in the videosequence. We developed the new algorithm of large 3D motion estimation by using the optical flow and the 3D model Candide. In the algorithm the prediction of 3D motion parameters which runs in the feedback loop with multiple iterations is applied. The prediction does not need the synthesized frames but only the frames of the input videosequence. Also the designed algorithm does not need continuous extraction of the feature points in the frames of the input videosequence, because they are given by the selected feature vertices of the model Candide.

The achieved experimental results show that the prediction of 3D motion parameters by the iteration process increases considerable the accuracy of estimation above all large 3D motion. Thereby the estimation error decreases including its small accumulation in long videosequences. Further achieved experimental results show the saturation state for 3 and more iterations when no increase of the accuracy of estimated 3D motion parameters occurs. Finally, objective and subjective evaluations of the experimental results show that the developed algorithm of large 3D motion estimation of the human head is suitable for using in the model based video coding of the videosequences where very high compression is expected. The

model based video coding is the important component of the standard videocodec MPEG-4 SNHC [17] which allows the advanced communications between the cloned and virtual human heads.

Acknowledgements

The work was supported by the Scientific Grant Agency of the Ministry of Education and the Academy of Science of the Slovak republic under Grant No. 1/3133/06.

References

[1] MIHALIK, J. *Image Coding in Videocommunnications*. Mercury-Smekal ISBN-80-89061-47-8, Košice, 2001. (In Slovak).

[2] MIHALÍK, J. Adaptive hybrid coding of images. *Journal of Electrical*, vol. 44, no.3, 1993, p.85-89. (In Slovak).

[3] FORCHHEIMER, R., KROMANDER, T. Image coding – from waveforms to animation. *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. 37, no. 12, 1989, p.2008-2023.

[4] PEARSON, D. E. Development in model-based video coding. *Proc. IEEE*, vol.83, no.6, 1995, p. 892-906.

[5] ANTOSZCZYSZYN, P. M., HANAH, J. M., GRANT, P. M. A new approach to wire-frame tracking for semantic model-based coding moving image, Coding, Signal Processing: *Image Communication* 15, 2000, p. 567-580.

[6] ZHANG, L. Estimation of eye and mouth corner point position in a knowledge-based coding system. *Proc. SPIE*, vol. 2952, 1996, p.21-28.

[7] DAVIS, M., TUCERYAN, M. Coding of facial image sequences by model-based optical flow. In *Proceedings of the Inter. Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*. Rhodes (Greece), September 1997, p. 192-194.

[8] LI, H., ROIVAINEN, P., FORCHHEIMER, R. 3-D motion estimation in model-based facial image coding. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, June 1993, p. 545-555.

[9] MIHALÍK, J., MICHALČIN, V. 3D Motion tracking of human head. In *Proc. of the 13th International Czech-Slovak Scient. Conf. "Radioelektronika 2003*", ISBN 80-214-2383-8, Brno (Czech Republic), 6-7 May 2003, p. 111-114.

[10] DECARLO, D., METAXAS, D. The integration of optical flow and deformable models with applications to human face shape and motion estimation. In *Proc.IEEE Conf. Computer Vision and Pattern Recognition.* IEEE CS Press, Los Alamitos (Calif.), 1996, p.231-238.

[11] HUANG, T. S., REDDY, S., AIZAWA, K. Human facial motion analysis and synthesis for video compression. In *SPIE Symposium on Visual Comm. and Image Proc.* Boston (MA, USA), November 1991, p. 234-241.

[12] PRAT, W. K. *Digital Image Processing.* John Wiley & Sons. New York, 1978.

[13] MIHALÍK, J., MICHALČIN, V.: 3D motion estimation and texturing of human head model. *Radioengineering*, vol. 13, no. 1, 2004, ISSN 1210-2512, p. 26-31.

[14] AHLBERG, J. Candide-3: *An Updated Parameterised Face. Rep. No. LiTH-ISY-R-2326*, January 2001.

[15] MICHALČIN, V. Calibration of 3D wire frame model of human head. In *Proc. of the IIIth Doctoral conference FEI TU*, Košice 2003, p. 65-66.

[16] MIHALÍK, J., MICHALČIN, V. Texturing of surface of 3D human head model. *Radioengineering,* ISSN1210-2512, Vol.13, No.4, 2004, p. 44- 47.

[17] The special issue of the *IEEE Trans.on Circuits and Systems for Video Technology on MPEG-4 SNHC*, July 2004.

About Authors...

Ján MIHALÍK graduated from the Technical University in Bratislava in 1976. Since 1979 he joined the Faculty of Electrical Engineering and Informatics of theTechnical University of Košice, where he received his PhD degree in Radio Electronics in 1985. Currently, he is Full Professor of Electronics and Telecommunications and the head of the Laboratory of Digital Image Processing and Videocommunications at the Department of Electronics and Multimedia Telecommunications. His research interests include information theory, image and video coding, digital image and video processing and multimedia videocommunications.

Viktor MICHALČIN was born on 1976 in Ukraine. He received the Ing degree from the Technical University of Košice in 2000. He is a PhD student at the Department of Electronics and Multimedia Telecomunications of the Technical University, Košice. His research is focused on model based and very low bit rate video coding. Currently he is working as a developer of VRVS/EVO videoconferencing system in Caltech-VRVS-SK team at the University of P. J. Šafarik in Košice.